Web-based Supplementary Materials for "New Semiparametric Method for Predicting High-Cost Patients"

Adam Maidman and Lan Wang
School of Statistics, University of Minnesota, Minneapolis, MN 55455, U.S.A.

Abstract

The regularity conditions for and proofs of (3.3) and (3.4) from the main article are given in Appendix A of this supplementary material. Additional numerical results from Section 4 are contained in Appendix B. Appendix C displays the prediction intervals for next-year expenditure in Panel 3 of MEPS. Lastly, a demonstration of the `R` package `plaqr` with sample code is provided in Appendix D.

Web Appendix A: Regularity conditions and proof of (3.4)

We can write $Y_i = V_i'\beta + \sum_{j=1}^q g_j(Z_{ij}) + \epsilon_i$, where the ϵ_i are independent and satisfy the constraint $P(\epsilon_i \leq 0|X_i) = \tau$, where $X_i = (V_i', Z_i')'$ with $V_i = (V_{i1}, \ldots, V_{ip})'$ and $Z_i = (Z_{i1}, \ldots, Z_{iq})'$.

Definition 0.1 *Let $r \equiv m + v$, where m is a positive integer and $v \in (0, 1]$. Define $\mathcal{H}_r$ as the collection of functions $h(\cdot)$ on $[0, 1]$ whose mth derivative $h^{(m)}(\cdot)$ satisfies the Hölder condition of order v. That is, for any $h(\cdot) \in \mathcal{H}_r$, there exists some positive constant C such that*

$$\left|h^{(m)}(z') - h^{(m)}(z)\right| \leq C|z' - z|^v, \quad \forall\ 0 \leq z', z \leq 1.$$

Definition 0.2 *Given $Z = (Z_1, \ldots, Z_q)'$, the function $g(Z)$ is said to belong to the class of functions $\mathcal{G}$ if it has the representation $g(Z) = \alpha + \sum_{k=1}^q g_k(Z_k)$, $\alpha \in \mathcal{R}$, $g_k \in \mathcal{H}_r$ and $E[g_k(Z_k)] = 0$.*

Let $h_j^*(\cdot) = \underset{h_j(\cdot)\in\mathcal{G}}{\arg\inf} \sum_{i=1}^n E\left[f_i(0)(x_{ij} - h_j(Z_i))^2\right]$, where $f_i(\cdot)$ is the probability density function of ϵ_i given X_i. Let $m_j(Z) = E\left[x_{ij}|Z_i = Z\right]$, then it can be shown that $h_j^*(\cdot)$ is the weighted projection of $m_j(\cdot)$ into $\mathcal{G}$ under the L_2 norm, where the weights $f_i(0)$ are included to account for the possibly heterogeneous errors. Define $\delta_{ij} \equiv X_{ij} - h_j^*(Z_i)$. Let V be the $n \times p$ matrix of the linear covariates. Let H be the $n \times q$ matrix with the (i, j)th element

$H_{ij} = h_j^*(Z_i)$, and write $V = H + \Delta$.

The following conditions are imposed for deriving the properties stated in Section 3.3. These conditions are similar to those in Sherwood and Wang (2016).

(C1) (Conditions on the random error) The random error ϵ_i has the conditional distribution function F_i and continuous conditional density function f_i . The f_i are uniformly bounded away from 0 and infinity in a neighborhood of zero and its first derivative f_i' has a uniform upper bound in a neighborhood of zero, for $1 \leq i \leq n$.

(C2) (Conditions on the covariates) There exist positive constants M_1 and M_2 such that $|V_{ij}| \leq M_1$, $\forall\ 1 \leq i \leq n$, $1 \leq j \leq p$ and $\mathrm{E}[\delta_{ij}^4] \leq M_2$, $\forall\ 1 \leq i \leq n$, $1 \leq j \leq q$. There exist finite positive constants C_1 and C_2 such that with probability one

$$C_1 \leq \lambda_{max}\left(n^{-1}VV'\right) \leq C_2, \quad C_1 \leq \lambda_{max}\left(n^{-1}\Delta\Delta'\right) \leq C_2.$$

(C3) (Condition on the non-linear functions) For $r = m + v > 1.5$, $g_0 \in \mathcal{G}$.

(C4) (Condition on the B-Spline basis) The dimension of the spline basis k_n satisfies $k_n \approx n^{1/(2r+1)}$. and $n^{-1}k_n^3 = o(1)$.

In the following, we will derive (3.4). In Sherwood and Wang (2016), it was proved that $||\hat{\beta} - \beta|| = O_p(n^{-1/2})$. We will first prove (3.5), which strengthens the result of Sherwood and Wang (2016) for estimating the nonparametric components to uniform convergence. To facilitate the proof, we will make use of the theoretically centered B-spline basis functions (e.g., Xue and Yang (2006)). More specifically, we consider the B-spline basis functions $b_j(\cdot)$ in Section 2 and let $B_j(z_{ik}) = b_{j+1}(Z_{ik}) - \frac{E[b_{j+1}(Z_{ik})]}{E[b_1(Z_{ik})]} b_1(Z_{ik})$ for $j = 1, \ldots, k_n + l$. Then $E(B_j(Z_{ik})) = 0$. For a given covariate Z_{ik}, let $\mathbf{w}(Z_{ik}) = (B_1(Z_{ik}), ..., B_{k_n+l}(Z_{ik}))'$ be the vector of basis functions, and $\mathbf{W}(Z_i)$ denote the J_n-dimensional vector $\left(k_n^{-1/2}, \mathbf{w}(Z_{i1})', ..., \mathbf{w}(Z_{iq})'\right)'$, where $J_n = q(k_n + l) + 1$.

By the result of Schumaker (1981, p. 227), there exists a vector $\boldsymbol{\gamma}_0 \in \mathcal{R}^{J_n}$ and a positive constant C_0, such that $\sup_{t \in [0,1]^d} |\sum_{k=1}^q g_k(\mathbf{t}) - \mathbf{W}(\mathbf{t})'\boldsymbol{\gamma}_0| \leq C_0 k_n^{-r}$. Let

$$(\hat{\mathbf{c}}_\mathbf{1}, \hat{\boldsymbol{\gamma}}) = \underset{(\mathbf{c}_1, \boldsymbol{\gamma})}{\operatorname{argmin}}\, n^{-1} \sum_{i=1}^n \rho_\tau\big(Y_i - V_i'\mathbf{c}_1 - \mathbf{W}(Z_i)'\boldsymbol{\gamma}\big). \tag{0.1}$$

We write $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_d')'$, where $\gamma_0 \in \mathcal{R}$, $\boldsymbol{\gamma}_j \in \mathcal{R}^{k_n+l}$, $j = 1, \ldots, d$; and we write $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1', \ldots, \hat{\boldsymbol{\gamma}}_d')'$ the same fashion. Let $\tilde{g}_j(Z_{ij}) = w(Z_{ij})'\hat{\boldsymbol{\gamma}}_j$ be the estimator of g_j, $j = 1, ..., q$. Let $\tilde{g}(Z_i) = \mathbf{W}(Z_i)'\hat{\boldsymbol{\gamma}} = \tilde{g}_0 + \sum_{j=1}^q \tilde{g}_j(Z_{ij})$; and $\hat{g}(Z_i) = \sum_{j=1}^q \widehat{g}_j(Z_{ij})$. It can be derived that $\hat{g} = \tilde{g}$.

$$\begin{aligned}
\sup_z \Big| \sum_{j=1}^q \widehat{g}_j(z) - \sum_{j=1}^q g_j(z) \Big| &= \sup_z \Big| \tilde{g}(z) - \sum_{j=1}^q g_j(z) \Big| \\
&= \sup_z \Big| \mathbf{W}(z)'(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \Big|
\end{aligned}$$

$$\leq \sup_z ||\mathbf{W}(z)|| \cdot ||\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0||.$$

Let $B_n = \text{diag}(f_1(0), \ldots, f_n(0))$ be the $n \times n$ diagonal matrix; $W = (\mathbf{W}(Z_1), ..., \mathbf{W}(Z_n))' \in \mathbb{R}^{n \times J_n}$, and $W_B^2 = W' B_n W \in \mathbb{R}^{J_n \times J_n}$. It follows from Sherwood and Wang (2016) that $||W_B(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)|| = O_p(\sqrt{k_n})|$. Hence,

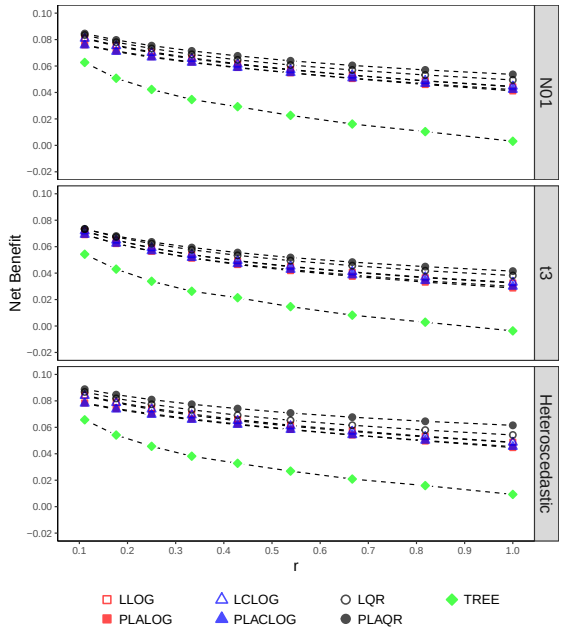
$$\begin{aligned} ||\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0|| &= ||W_B^{-1} W_B(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)|| \\ &\leq \sqrt{k_n n^{-1}} O_p(k_n) = O_p(k_n^{3/2} n^{-1/2}). \end{aligned}$$

In our setting, $\sup_z ||\mathbf{W}(z)|| = O_p(1)$. Thus $\sup_z \left| \sum_{j=1}^q \widehat{g}_j(z) - \sum_{j=1}^q g_j(z) \right| = O_p(k_n^{3/2} n^{-1/2}) = o_p(1)$. Hence (3.5) is verified. We thus have

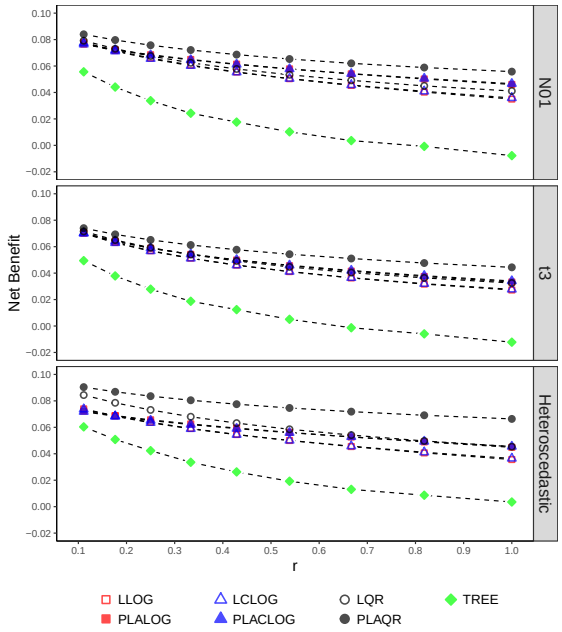
$$\begin{aligned} & \text{sign}[\hat{Q}_{Y^*|X^*}(0.5) - c] \\ = \; & \text{sign}[Q_{Y^*|X^*}(0.5) - c + V_i^{*'}(\hat{\beta} - \beta) + \sum_{j=1}^q (\hat{g}_j(Z_{ij}) - g_j(Z_{ij}))] \\ = \; & \text{sign}[Q_{Y^*|X^*}(0.5) - c] + o_p(1) \end{aligned}$$

since $\hat{\beta}$ and $\hat{g}_j$, $j = 1, \ldots, q$, are estimated on the training data and are independent of (Y^*, X^*). □

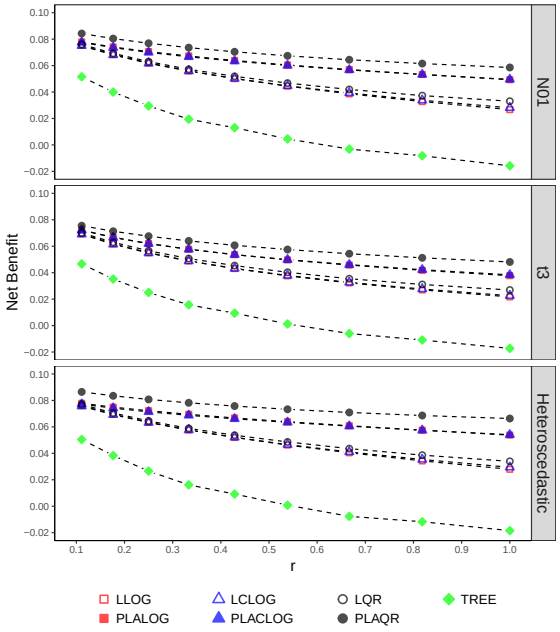
Web Appendix B: Additional numerical results from Section 4.1



(a) b=1



(b) b=2

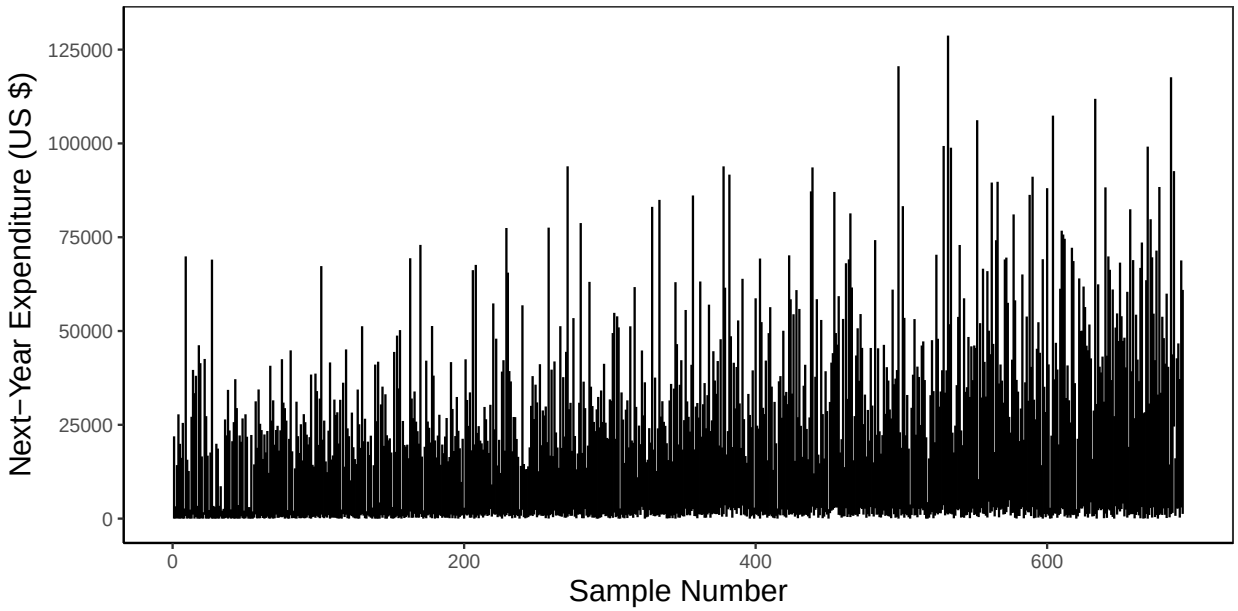


(c) b=3

Web Figure 1: Decision curves for the LLOG, PLALOG, LCLOG, PLACLOG, LQR, TREE, and PLAQR procedures for simulations with standard normal errors, t_3 errors, and heteroscedastic errors when $b = 1, 2, 3$. All standard errors are less than 2.7×10^{-4}

.

Web Appendix C: Prediction Intervals for Next-Year Expenditure



Web Figure 2: 90% prediction intervals for next-year expenditure in Panel 3 of MEPS.

Web Appendix D: Sample code for using `plaqr` package

```
library(plaqr)
set.seed(4)

n = 1000

### Generate the covariates
x1 <- rnorm(n); x2 <- rnorm(n)
z1 <- runif(n); z2 <- runif(n, -1,1)

### Generate the response
y <- exp( x1 + x2 + sin(2*pi*z1) + z2^3 + rnorm(n) )

### Customize the settings for the spline basis functions for z1 and z2
splinesettings <- vector("list", 2)
splinesettings[[2]]$degree <- 4
splinesettings[[2]]$Boundary.knots <- c(-1,1)

### Estimate the transformation parameter
trans <- transform_plaqr(y ~ x1 + x2, ~ z1 + z2, tau=.5,
```

```
           splinesettings=splinesettings, lambda=seq(0,3,by=.05))
trans$parameter
### Save the transformed response
newy <- trans$Y

### Fit the model
fit <- plaqr(newy ~ x1 + x2, ~ z1 + z2, tau=.5,
           splinesettings=splinesettings)

### Plot the nonlinear effects
plot( nonlinEffect(fit) )

### Make prediction intervals
newdata <- data.frame( x1=c(-1,1), x2=c(0,3),
        z1=c(.2, .6), z2=c(-.5,-.75) )
intervals <- predictInt( fit, newdata=newdata )

### Transform the intervals back to original scale
trans_parameter(intervals, trans$parameter, inverse=TRUE)
```


Acknowledgements

We thank the Co-Editor, the associate editor, and the anonymous referees for their helpful comments which helped us improve the paper significantly. The research was partially supported by DMS-1712706 and a grant from the U.S. Department of Veterans Affairs.

References

Maidman, A. (2017). *plaqr: Partially Linear Additive Quantile Regression*. R package version 2.0 (available from `http://CRAN.R-project.org/package=plaqr`).

Schumaker, L. (1981). *Spline functions: basic theory*. John Wiley&Sons, New York, NY, USA.

Sherwood, B. and Wang, L. (2016). Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics* **44,** 288–317.

Xue, L. and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica* **16,** 1423.